NAIIS 2018
NIGERIA HIV/AIDS INDICATOR AND IMPACT SURVEY
...getting to zero

NIGERIA HIV/AIDS INDICATOR AND IMPACT SURVEY

SAMPLING AND WEIGHTING TECHNICAL REPORT

NOVEMBER 2020

PARTNERS

FEDERAL MINISTRY OF HEALTH
NATIONAL AGENCY FOR THE CONTROL OF AIDS NACA
NATIONAL POPULATION COMMISSION
nbs NATIONAL BUREAU OF STATISTICS
CDC CENTERS FOR DISEASE CONTROL AND PREVENTION
The Global Fund
UNAIDS
World Health Organization
UNIVERSITY of MARYLAND SCHOOL OF MEDICINE
INSTITUTE OF HUMAN VIROLOGY
UNIVERSITY of WASHINGTON
ICF
AFENET African Field Epidemiology Network A healthier Africa

This project is supported by the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) through CDC under the terms of cooperative agreement #U2GGH002108.
The contents of this document do not necessarily represent the official position of the funding agencies.

NIGERIA HIV/AIDS INDICATOR AND IMPACT SURVEY (NAIIS) 2018
SAMPLING AND WEIGHTING TECHNICAL REPORT

NAIIS 2018 Collaborating Institutions

Federal Ministry of Health, Nigeria (FMoH)
National Agency for the Control of AIDS, Nigeria (NACA)
National Population Commission, Nigeria (NPopC)
National Bureau of Statistics, Nigeria (NBS)
The United States Centers for Disease Control and Prevention (CDC)
The Global Fund to Fight AIDS, Tuberculosis and Malaria (GF)
Center for International Health, Education, and Biosecurity (Ciheb) at the University of Maryland, Baltimore (UMB)
ICF International
African Field Epidemiology Network (AFENET)
University of Washington (UW)
The Joint United Nations Programme on HIV and AIDS (UNAIDS)
World Health Organization (WHO)
United Nations Children's Fund (UNICEF)

Donor Support and Disclaimer

This project is supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC) under the cooperative agreement #U2GGH002108 to the University of Maryland, Baltimore (UMB) and by the Global Fund to Fight AIDS, Tuberculosis and Malaria through the National Agency for the Control of AIDS (NACA), Nigeria, under the contract #NGA-H-NACA to UMB. The findings and conclusions of this report are those of the authors and do not necessarily represent the official position of the funding agencies.

Suggested Citation

Federal Ministry of Health, Nigeria. Nigeria HIV/AIDS Indicator and Impact Survey (NAIIS) 2018: Sampling and Weighting Technical Report. Abuja, Nigeria. November 2020.

Access This Report Online

www.nigerianstat.gov.ng/nada/index.php/catalog/68

Contact Information

Federal Ministry of Health
New Federal Secretariat Complex
Phase 3
Ahmadu Bello Way
PMB 083 Garki, Abuja
Phone: +234 9 5238362
Email: info@nigeria.gov.ng
Website: www.fmh.gov.ng

National Agency for the Control of AIDS
No.3 Ziguinchor Street
Wuse Zone 4, Abuja
Phone: +234 9 4613726
Email: info@naca.gov.ng
Website: www.naca.gov.ng

Contents

List of Tables and Figures

List of Abbreviations

CHAID: Chi-square Automatic Interaction Detector

CI: Confidence Interval

CMOS: Cumulative Measure of Size

EA: Enumeration Area

FCT: Federal Capital Territory

LASSO: Least Absolute Shrinkage and Selection Operator

LGA: Local Government Area

MOS: Measure of Size

NAIIS: Nigeria HIV/AIDS Indicator and Impact Survey

NPopC: Nigeria National Population Commission

NSUM: Network Scale-Up Method

PPS: Probability Proportional to Size

PSU: Primary Sampling Unit

RS: Random Start

SI: Sampling Interval

SRS: Simple Random Sample

1. Introduction

The Nigeria HIV/AIDS Indicator and Impact Survey (NAIIS) is a two-stage stratified cluster sample survey designed to assess the prevalence of key Human Immunodeficiency Virus (HIV)-related health indicators. Data collection for NAIIS occurred between July and December 2018 and had a sample size of 83,909 households and 383,574 individuals (aged 0 to 64 years) across 36 states and the Federal Capital Territory (FCT). The purpose of this report is to document the procedures used to select the households and individuals for the study and the subsequent weighting of the respondent sample.

1.1 Overview of Sample Design

NAIIS sampled the population using a two-stage cluster sampling technique, selecting enumeration areas (EAs) followed by households. The first-stage sampling units (also referred to as the "primary sampling units" or PSUs) were stratified by the 36 states and the FCT. The sample size was calculated to provide a representative national estimate of HIV incidence and HIV prevalence among adults aged 15-64 years and was also calculated to provide HIV prevalence estimates at the state level. One-quarter of the households were randomly selected for inclusion of children, which was designed to provide a representative national estimate of pediatric HIV prevalence.

The second-stage sampling units were selected from lists of dwelling units/households compiled by trained staff for each of the sampled PSUs. At the request of Lagos State, the NAIIS sample design was adjusted to oversample Lagos State to obtain stable estimates of HIV prevalence in 20 Local Government Areas (LGAs). Within the sampled households, all eligible adults aged 15 to 64 years were included in the study sample for data collection. All eligible children aged 0-14 years in a randomly designated subset of the selected households were included in the sample.

Details of the sample design utilized for NAIIS are provided in Section 2.

1.2 Overview of Weighting Process

The main purpose of the survey weights calculated for NAIIS was to 1) account for unequal selection probabilities at different stages of sampling, 2) adjust for nonresponse at different stages of data collection, 3) reduce the variability of the weights using a weight trimming procedure, and 4) calibrate the weights to the 2018 population projections using data from the Nigeria National Population Commission (NPopC).

Taking into consideration the objectives above, the process of calculating the weights started by calculating the design weights that account for the selection probabilities of the different sampling units in different sampling stages. The design weights were adjusted to account for nonresponse that happened on the PSU and household levels. When weights are calculated for individuals, such as adults or adolescents, the weights were adjusted for individual-level nonresponse to the survey questionnaire. When weights are calculated for measurements, such as blood draws for HIV, the weights were adjusted for nonresponse to the test. All weights were trimmed, where outliers were capped at a maximum value. Finally, all weights were calibrated based on the percentage or total distributions of the projected population.

Technical details of the weighting procedures employed in NAIIS are provided in Section 3.

2. Sample Design

2.1 Population of Inference

The population of inference for NAIIS was comprised of the de facto household population. The *de facto* population was comprised of individuals who were present in households, i.e., slept in the household, on the night prior to the household interview. In contrast, the *de jure* population is comprised of individuals who are usual residents of the household, irrespective of whether they slept in the household on the night prior to the household interview.

2.2 Precision Specifications and Assumptions

The following specifications and assumptions were used to develop the sample design for NAIIS.

2.2.1 Specifications

- The relative standard error of the national estimate of HIV incidence among persons aged 15-64 was set at ~30%.
- The 95% confidence bounds (also known as the *margin of error*) were used for the estimated VLS rate among HIV-positive persons aged 15-64 in each of the 37 strata (states) calculated at ~10%.

2.2.2 Assumptions

- An overall HIV prevalence rate of 3.4% that varied by state.
- An annual HIV incidence rate for adults aged 15-64 of 0.49%.
- A mean duration of recent HIV infection (MDRI) of 130 days, yielding an annualization rate of 365/130= 2.8077. Hence, the estimated HIV incidence rate for MDRI = 130 days was *Pm*=0.0060/2.8077=0.0021 (0.21%).
- The VLS rate among HIV-positive adults aged 15-49 in each state h of *Pvh* = 50%. This was a conservative assumption because it overstated the actual variance of the VLS rate.
- An intra-cluster correlation (ICC) of 0.02 for both prevalence and incidence. The ICC provided an average measure of the homogeneity of responses within the first-stage sampling units.
- An occupancy rate of 100% was used for sampled dwelling units. Note that this was not included in the calculation of the overall survey response rate but does determine the initial numbers of dwelling units to be sampled.
- An overall household response rate of 90.6% was witnessed among the occupied dwelling units.[1]
- The average number of persons aged 15-64 in a household was 2.47.[1]
- The percentage of persons in households who were aged 0-14 was 45.7%.[1]
- The percentage of persons in households who were aged 15-64 was 48.2%.[1]
- Among individuals aged 15-64 in eligible responding households, the biomarker response rate was 77.3%. This corresponded to an overall biomarker response rate of 63%. This was a conservative estimate derived from response rates in the 2012 National HIV & AIDS and Reproductive Health Survey (NARHS 2012).[1]
- Among children aged 0-14 in eligible responding households, the biomarker response rate was 63%.

[1] The assumed values of response rates and the number of participating persons per household was based on data from the 2013-14 Nigeria Demographic and Health Survey (DHS) NARHS 2012.

2.3 Selection of the Primary Sampling Units (PSUs)

2.3.1 Definition of PSUs

The sampling frame consisted of 662,855 EAs containing 28,900,478 million households and 140,431,798 million persons. There were an average number of households and persons per EA of 44 and 212, respectively. The EAs were mutually exclusive (non-overlapping). This ensured that all households and residents had a chance of being included in the survey. Given the variability in household size across Nigeria (range between 4.0 to 5.7 individuals per household), state differences in household size based on the 2006 Census were considered when calculating the number of EAs or PSUs to be selected in each state.

Since each of the states and FCT was a reporting domain for HIV prevalence, the number of clusters allocated to each state and the FCT was sampled on a state-by-state basis. All the EAs for each state were obtained and arranged in their geographic order and the projected counts of households of each EA was attached as a Measure of Size (MOS) of the EA. Within each state, a probability proportional to size (PPS) sample was selected, where a Cumulative Measure of Size (CMOS) was obtained by successive addition of the individual MOS in the ordered list. The total projected count of households of the state (which is the same as the Last CMOS) was divided by the number of EAs (n) allocated to the state for the survey to obtain a Sampling Interval (SI) for the Probability Proportion to Size systematic sample used for sampling the PSUs in each state and the FCT. Thereafter, a random number between 1 and the SI was generated using the computer (Excel random number generation). This randomly generated number serves as the Random Start (RS) for the sampling. The SI was added to the obtained RS (n-1) successive times to generate "n" sequential numbers as follows: RS, RS+SI, RS+2SI, RS+3SI, ..., RS+(n-1)SI.

The EAs with the CMOS range that contains each of the n sequence numbers as generated above were selected for NAIIS.

2.3.2 Selection of the PSU Sample

A stratified sample of 4,035 EAs was selected from the sampling frame. These EAs were used as the PSUs that served as the survey clusters. The 37 strata specified for sampling were the 36 states and the FCT. The EA samples were selected systematically and with probabilities proportionate to a measure of size (MOS) equal to the 2018 projected number of households in the EA based on the 2008 census. Prior to selection, the EAs were sorted by type of EA, including urban/rural and other geographic variables in the frame. The sorting of the EAs prior to sample selection induces an implicit geographic stratification. Within each stratum, a sample of EAs was systematically selected with PPS selection.

2.4 Selection of Households

The selection of households for NAIIS involved the following steps: (1) listing the dwelling units/ households within the sampled EAs, (2) assigning eligibility codes to the listed dwelling unit/household records, (3) selecting the samples of dwelling units/households, and (4) designating a subsample of households for child data collection.

For both sampling and analysis purposes, a household is defined as a group of persons who normally live and eat together. These people may or may not be related by blood but make common provision for food or other essentials for living and they have only one person whom they all regarded as the head of the household. Households were eligible for participation in this survey if they were within the pre-defined EA and were randomly selected for inclusion in the survey.

2.4.1 Listing

Individuals that had experience in mapping and listing of households, including those that had participated in mapping and listing in projects such as the Nigeria Demographic Health Survey (NDHS) and the National HIV/AIDS and Reproductive Health Survey (NAHRS) were identified across the states and recruited for mapping NAIIS. There was an initial central level training of 6 zonal coordinators and 36 state and FCT coordinators in Abuja. The zonal and state coordinators went back to the respective states in their zones to train the mapping and listing supervisors and Mapper/Listers.

The following activities were conducted by the Mapper/Listers after training:

- Obtained the physical EA Map of the sampled EAs from the National Population Commission office.
- Approached the local community leaders of the sampled EAs for permission to map and list households after explaining their mission.
- Beginning with the "starting point" on the EA map, the mappers identified current structures within the boundaries of the sampled EAs.
- Numbered all the buildings/structures within the EA as NAIIS/Cluster Number/ Building number. Example: NAIIS/3205/19 on the walls.
- Explained their mission to occupants of residential buildings.
- Explained the concept of "Household" to the occupant of all the residential buildings.
- Identified the Household(s) occupying the respective buildings and listed each household indicating the address or name of the owner of the building, and the name of the Head of Household. This information was captured using an electronic data capture tablet and streamed in real-time to a central server.
- Sketched the EA on A3 paper indicating all current structures and landmarks within the EA boundaries.
- Described in detail on the back of the Sketched Map, the location of the EA, indicating how a visitor could travel to the EA from a central popular Motor Park in the state capital. The description detailed how and where to reach the EA using public transport (bus, taxi, motorcycle, trekking), as well as the names and contact information of individuals that helped during the listing.

2.4.2 Households Selection

Within each EA, a random sample of households was selected from households listed during the listing process. Using a systematic sampling procedure, a total of 28 households were sampled from each cluster in all states except for Lagos state where 9 households were sampled per cluster as Secondary Sampling Unit (SSU). This sample of households served as the frame (Frame 2) for a sub-sample of households to be selected for the Hepatitis B&C and/or Pediatric testing.

By applying a Simple Random Sample (SRS) procedure, a sub-sample of the above-sampled households (Frame 2) was selected to yield the number of households allocated to participate in the Hepatitis B&C test in each state. In the same manner, another sub-sample was obtained using SRS procedure from Frame 2 and tagged as households that would participate in the Pediatrics test.

The lists of the 28 or 9 households sampled in each cluster with their identification particulars including locality name, cluster number, building number, address of building, and name of head of household were extracted and printed for the field data collection team to trace and interview the selected households. The lists also indicated if the sampled household would participate only in the general survey or would also participate in either or both the Hepatitis and Pediatric samples.

2.5 Selection of Individuals

The selection of individuals for NAIIS involved the following steps: (1) compiling a list of all individuals known to reside in the household or who slept in the household during the night prior to data collection; (2) identifying those rostered individuals who were eligible for data collection; and (3) selecting for the study those individuals meeting the age and residency requirements of the study. However, only those individuals who slept in the household the night before the household interview, i.e., the *de facto* population, were retained for subsequent weighting and analysis.

In all households, all adults aged 15-64 years were eligible to complete an adult interview and to be tested. In addition, one adult aged 18-54 years selected at random from each household was eligible to complete the network scale-up method (NSUM) module. In households selected for the pediatric subsample, all adolescents aged 10-14 years were eligible to complete an interview, and all children aged 0-14 years were eligible for blood testing. All adults aged 15-64 years who tested positive for HIV were eligible for Hepatitis B and Hepatitis C testing. In households selected for Hepatitis B and Hepatitis C testing (Hepatitis subsample), one adult aged 15-64 years was selected at random from each household to be tested for Hepatitis B and Hepatitis C regardless of their HIV status. Table 1 provides the allocation of the clusters and households selected by state for each type of sample.

Table 1: Distribution of sampled enumeration areas and households by state

State	Total clusters sampled for the survey	Number of households sampled for the survey	Number of households sampled for pediatrics test	Number of households sampled for hepatitis B and C test
Abia	101	2,828	601	233
Adamawa	88	2,464	582	265
Akwa Ibom	104	2,912	846	344
Anambra	100	2,800	875	347
Bauchi	87	2,436	845	411
Bayelsa	100	2,800	358	143
Benue	89	2,492	795	357
Borno	92	2,576	799	365
Cross river	106	2,968	641	242
Delta	103	2,884	888	356
Ebonyi	98	2,744	446	178
Edo	103	2,884	697	264
Ekiti	99	2,772	494	203
Enugu	105	2,940	717	275
FCT1	105	2,940	309	215
Gombe	86	2,408	424	203
Imo	101	2,828	828	342
Jigawa	89	2,492	811	354
Kaduna	89	2,492	1,133	513
Kano	82	2,296	1,615	817

Table 1: Distribution of sampled enumeration areas and households by state (continued)				
State	**Total clusters sampled for the survey**	**Number of households sampled for the survey**	**Number of households sampled for pediatrics test**	**Number of households sampled for hepatitis B and C test**
Katsina	87	2,436	1,061	490
Kebbi	83	2,324	569	276
Kogi	92	2,576	637	277
Kwara	95	2,660	470	191
Lagos	600	5,400	2,215	777
Nasarawa	89	2,492	349	204
Niger	89	2,492	735	337
Ogun	112	3,136	877	324
Ondo	105	2,940	756	291
Osun	102	2,856	727	304
Oyo	107	2,996	1,249	491
Plateau	90	2,520	602	261
Rivers	103	2,884	1,125	455
Sokoto	88	2,464	685	312
Taraba	91	2,548	435	201
Yobe	89	2,492	433	206
Zamfara	86	2408	591	281
Total	4,035	101,580	28,220	12,105
[1] FCT: Federal Capital Territory				

3. Weighting and Estimation

In general, the purpose of weighting survey data from a complex sample design is to (1) compensate for variable probabilities of selection, (2) account for differential nonresponse rates within relevant subsets of the sample, and (3) adjust for possible under coverage of certain population groups. Weighting is accomplished by assigning an appropriate sampling weight to each responding sampled unit (e.g., a household or person), and using that weight to calculate weighted estimates from the sample. The critical component of the sampling weight is the base weight which is defined to be the reciprocal of the probability of including a household or person in the sample. The base weights are used to inflate the responses of the sampled units to population levels and are generally unbiased (or consistent) if there is no nonresponse or noncoverage in the sample (e.g., see Kish, 1965, p. 67). When nonresponse or noncoverage occurs in the survey, weighting adjustments are applied to the base weights to compensate for both types of sample omissions.

Nonresponse is unavoidable in virtually all surveys of human populations. For NAIIS, nonresponse can occur at different stages of data collection, for example, (1) before the enumeration of individuals in the household, (2) after household enumeration and selection of persons but before completion of the individual interview, and (3) after completion of the interview but before collection of a usable blood sample. The procedures used to compensate for nonresponse at each of the relevant stages of data collection are described in Section 3.4.

Noncoverage arises when some members of the survey population have no chance of being selected for the sample. For example, noncoverage can occur if the field operations fail to enumerate all dwelling units during the listing process, or if certain household members are omitted from the household rosters. To compensate for such omissions, the poststratification procedures described in Sections 3.4.3.4 and 3.4.4.4 are used to calibrate the weighted sample counts to available population projections.

3.1 Overview of Survey Weights

The following six survey weights were calculated for the NAIIS data:

1. **Household survey weight (hhwgt):** weight for the household interviews.
2. **Adult interview weight (adwgt):** individual weight for adults aged 15-64 years in all households.
3. **Adolescent interview weight (adowgt):** individual weight for adolescents aged 10-14 years interviewed in the pediatric subsample - a subsample of households selected for the adolescent interviews and blood tests for children 0-14 years.
4. **NSUM weight (nswgt):** weight for the NSUM questions asked for one adult aged 18-54 years per household.
5. **Blood draw weight (bdwgt):** weight for blood tests for adults aged 15-64 years in all households and children aged 0-14 years in the pediatric subsample.
6. **Hepatitis weight (hepwgt):** weight for Hepatitis B/C tests for adults aged 15-64 years who tested HIV-positive and adults aged 15-64 years who were selected in the Hepatitis subsample - one adult aged 15-64 years per household from a subsample of households selected for Hepatitis tests.

3.2 Preparation for Weighting

The following data files were used during the weight calculations:

NAIIS_MERGED_04102019: a merged SAS dataset of the following SAS datasets:

- **NAHH_HOUSEHOLD:** a dataset that includes records for all selected households (HH) and data collected in the HH questionnaire for completed households.
- **NAHH_INDIVIDUAL:** a dataset that contains records of all individuals rostered in the completed households collected in the HH questionnaire.
- **NAIN_15_64:** Interview records for individuals aged 15-64 years collected from the adult questionnaire.
- **NAIN_0_14:** Interview records for individuals aged 10-14 years from the adolescent questionnaire and records from module 4 of the adult questionnaire about children aged 0-9 years.
- **NACN_INDIVIDUAL:** Counselling records for all participants that were counselled for testing.
- **BIOM_INDIVIDUAL:** Household biomarker records for all participants.
- **Naiis_lab_data_27_02_2019_si:** Satellite and Central lab biomarker records for all participants.

ALLSTATES_SELPROB: a CSV dataset that includes the selection probabilities for all the selected sampling clusters for the NAIIS. The file includes the selection probabilities for the different stages and the different subsamples.

3.3 Eligible and Completed Households and Individuals

For each weight, the calculation process starts by identifying the eligible units, households or individuals, and the completed cases, including completed questionnaires or valid blood measurements. Table 2 presents variables used to identify the eligible and completed households and individuals for each weight.

Table 2: Variables of eligibility and completion of each module/weight

Weight	Eligible units		Completed units
	Description	Variables (codes)	Variables (codes)
Household survey weight (hhwgt)	All selected households	AHRESULT (1,2,4,5,8)	AHRESULT (1)
Adult interview weight (adwgt)	All de-facto adults 15-64 in completed households	AGE (15:64) + AHSLEPT (1) + AHRESULT (1)	ARESULT (1)
Adolescent interview weight (adowgt)	All de-facto adolescents 10-14 in completed households in the pediatric subsample	AGE (10:14) + PEDIATRIC (1) + AHSLEPT (1) + AHRESULT (1)	ARESULT (1)
NSUM weight (nswgt)	One de-facto adult 18-54 per household in completed households	AGE (18:54) + ARESULT (1) + AHSLEPT (1) + M1101 (0:99)	M1101 (0:99)

Table 2: Variables of eligibility and completion of each module/weight (continued)			
Weight	**Eligible units**		**Completed units**
	Description	**Variables (codes)**	**Variables (codes)**
Blood draw weight (bdwgt)	All de-facto adults 15-64 who completed the adults questionnaire and all de-facto children 0-14 in completed households in the pediatric subsample	AGE (15:64) + AHSLEPT (1) + ARESULT (1) & AGE (0:14) + PEDIATRIC (1) + AHSLEPT (1) + AHRESULT (1)	Final_HIV_Status (1,2)
Hepatitis weight (hepwgt)	All de-facto adults 15-64 who tested positive for HIV and all de-facto adults 15-64 who were selected for Hepatitis testing in the Hepatitis subsample (one adult 15-64 per household)	AGE (15:64) + AHSLEPT (1) + Final_HIV_Status (1) & AGE (15:64) + ELGHEP (1) + AHSLEPT (1) + Final_HIV_Status (2) + HBRESULT (1,2)	HBRESULT (1,2)

3.4 Development of Weights

3.4.1 The Design Weight

Since the NAIIS sample is a two-stage stratified cluster sample stratified by states, the process of calculating the survey weights started by accounting for the different sampling probabilities, which were calculated separately for each sampling stage and each cluster. Let P_{1hi} be the first stage's sampling probability of the i^{th} cluster in stratum h, P_{2hi} is the second-stage's sampling probability of households within the i^{th} cluster and P_{hi} is the overall sampling probability of any households of the i^{th} cluster in stratum h. The probability of selection of the PSU i in stratum h in the sample, using the probability proportional to size method, is calculated as follows:

$$P_{1hi} = \frac{a_h N_{hi}}{\sum_{i=1}^{a_h} N_{hi}}$$

where a_h denote the number of clusters selected in stratum h, N_{hi} the number of households according to the sampling frame in the i^{th} cluster, and $\sum_{i=1}^{a_h} N_{hi}$ the total number of households in the stratum h. Now, let L_{hi} be the number of households listed during the household listing operation in cluster i in stratum h and let g_{hi} be the number of households selected in the same cluster. The second stage's selection probability for each household in the cluster is calculated as follows:

$$P_{2hi} = \frac{g_{hi}}{L_{hi}}$$

The overall selection probability of each household in cluster i of stratum h is, therefore, the product of the two stages of selection probabilities:

$$P_{hi} = P_{1hi} \times P_{2hi}$$

Therefore, the design weight for each household in cluster i of stratum h is the inverse of its overall selection probability as follows:

$$d_{hi} = \frac{1}{P_{hi}}$$

To prepare a final design weight so it can be used to calculate the final weights, d_{hi} was adjusted for cluster-level non-response to account for selected PSUs that were not completed due to flooding or security reasons. Let R_i identifies completed PSU *i*, where R_i=1 if PSU *i* was completed, and R_i = 0 if PSU *i* was not. Sampling design strata were used to form the nonresponse adjustment cells where a separate adjustment factor was calculated for each adjustment cell *c* as:

$$A_c^{PSU} = \sum_{i=1}^{a_c} d_{hi} \Big/ \sum_{i=1}^{a_c} R_i d_{hi}$$

where a_c is the number of sampled PSUs in adjustment cell *c*. State-level weighted response rates are presented in Table 3. For the completed PSUs, the nonresponse adjusted design weight for PSU *i* in stratum *h* was then computed as:

$$D_{hi} = A_c^{PSU} d_{hi}$$

The design weight D_{hi} is the base for all the survey weights explained in the following sections.

Table 3: Number of selected and dropped Primary Sampling Units (PSUs) and weighted response rates by state

State	Selected PSUs	Dropped PSUs	Reason	Weighted Response Rate ($1/A_c^{PSU}$)
Abia	101	0	N/A	N/A
Adamawa	88	4	Security	0.9372
Akwa Ibom	104	0	N/A	N/A
Anambra	100	0	N/A	N/A
Bauchi	87	0	N/A	N/A
Bayelsa	100	0	N/A	N/A
Benue	89	0	N/A	N/A
Borno	92	42	Security	0.8857
Cross River	106	0	N/A	N/A
Delta	103	0	N/A	N/A
Ebonyi	98	0	N/A	N/A
Edo	103	0	N/A	N/A
Ekiti	99	0	N/A	N/A
Enugu	105	0	N/A	N/A
FCT[1]	105	0	N/A	N/A
Gombe	86	0	N/A	N/A
Imo	101	0	N/A	N/A
Jigawa	89	1	Security	0.9901
Kaduna	89	10	Security	0.9091
Kano	82	0	N/A	N/A
Katsina	87	7	Security	0.9234
Kebbi	83	4	Security	0.9398

Table 3: Number of selected and dropped Primary Sampling Units (PSUs) and weighted response rates by state (continued)				
State	**Selected PSUs**	**Dropped PSUs**	**Reason**	**Weighted Response Rate ($1/A_c^{PSU}$)**
Kogi	92	0	N/A	N/A
Kwara	95	1	Flood	0.9804
Lagos	600	0	N/A	N/A
Nasarawa	89	1	Flood	0.9823
Niger	89	0	N/A	N/A
Ogun	112	0	N/A	N/A
Ondo	105	0	N/A	N/A
Osun	102	0	N/A	N/A
Oyo	107	0	N/A	N/A
Plateau	90	0	N/A	N/A
Rivers	103	0	N/A	N/A
Sokoto	88	9	Security	0.9234
Taraba	91	1	Security	0.9901
Yobe	89	19	Security	0.7663
Zamfara	86	40	Security	0.6793
Total	**4035**	**139**		
[1] FCT: Federal Capital Territory				

3.4.2 Household Survey Weights

The first step of calculating the household survey weight *hhwgt* was to adjust the design weight D_{hi} for household non-response to account for eligible non-respondents and units with unknown eligibility. Each sampled household was assigned to one of the following nine household interview result codes:

1. Completed
2. No household member at home or no competent respondent at home at time of visit
3. Entire household absent for extended period of time
4. Postponed
5. Refused
6. Dwelling vacant or address not a dwelling
7. Dwelling destroyed
8. Dwelling not found
9. Other

Table 4 shows the mapping of these 9 result codes to 4 response status groups: (1. Eligible respondents, 2. Eligible non-respondents, 3. Unknown eligibility and 4. Ineligible/out-of-scope).

Table 4: Number of selected households by response groups		
The household interview result	**Number of households**	**The response status group**
1. Completed	83909	1.Eligible respondents
2. No household member at home or no competent respondent at home at time of visit	1048	2.Eligible non-respondents
3. Entire household absent for extended period of time	5882	4.Ineligible/out-of-scope
4. Postponed	0	2.Eligible non-respondents
5. Refused	4388	2.Eligible non-respondents
6. Dwelling vacant or address not a dwelling	781	4.Ineligible/out-of-scope
7. Dwelling destroyed	32	4.Ineligible/out-of-scope
8. Dwelling not found	171	3.Unknown eligibility
9. Other	980	4.Ineligible/out-of-scope

The household nonresponse adjustment for eligible non-respondents and units with unknown eligibility was implemented in one step. Each household *j* was assigned to one of the 4 eligible response status groups. R_{hij} identifies respondent households *j* in PSU *i* in stratum *h*, where R_{hij}=1 if the household *j* was assigned to the first response group, and R_{hij} =0 if the household *j* was assigned to the second or third response status groups. Households in the fourth response group are ineligible for the survey and therefore are not part of this adjustment or the weight calculations in general. Rural/urban areas within States were used to form the nonresponse adjustment cells where a separate adjustment factor was calculated for each adjustment cell *c* as:

$$A_c^{HH} = \sum_{j=1}^{n_c} D_{hij} \Big/ \sum_{j=1}^{n_c} R_{hij} D_{hij}$$

where n_c is the number of sampled households in adjustment cell *c* not including the ineligible households assigned to response status group number. The adjustment factors A_c^{HH} are presented in Table 5. For the respondent households in response-status group 1, the nonresponse adjusted weight for household *j* in PSU *i* in stratum *h* was then computed as:

$$hhwgt_{hij} = A_c^{HH} D_{hij}$$

Table 5: Weighted sums of design weight for all and completed households and nonresponse adjustment factors by state and urban/rural				
		Weighted sums of design weight		**HH nonresponse adjustment factors**
State	**Urban/Rural**	**All households**	**Completed households**	
Abia	Urban	236147.5	209801.6	1.126
Abia	Rural	525278.6	485139.0	1.083
Adamawa	Urban	233785.7	205264.5	1.139
Adamawa	Rural	634623.4	579154.9	1.096
Akwa Ibom	Urban	177676.8	143228.2	1.241

Table 5: Weighted sums of design weight for all and completed households and nonresponse adjustment factors by state and urban/rural (continued)				
		Weighted sums of design weight		HH nonresponse adjustment factors
State	Urban/Rural	All households	Completed households	
Akwa Ibom	Rural	958412.9	847731.6	1.131
Anambra	Urban	893461.8	744632.5	1.200
Anambra	Rural	165457.8	138511.4	1.195
Bauchi	Urban	138940.4	118541.9	1.172
Bauchi	Rural	822532.5	759348.8	1.083
Bayelsa	Urban	135486.5	118273.4	1.146
Bayelsa	Rural	414595.0	349830.1	1.185
Benue	Urban	173633.2	158757.5	1.094
Benue	Rural	1407875.1	1264812.2	1.113
Borno	Urban	639693.6	478823.6	1.336
Borno	Rural	293052.3	246593.9	1.188
Cross River	Urban	140338.3	113589.5	1.235
Cross River	Rural	653326.7	535618.6	1.220
Delta	Urban	464223.5	401253.0	1.157
Delta	Rural	739794.6	643040.7	1.150
Ebonyi	Urban	176065.7	166816.8	1.055
Ebonyi	Rural	749016.2	712307.3	1.052
Edo	Urban	507994.1	440779.4	1.152
Edo	Rural	329934.4	300659.8	1.097
Ekiti	Urban	588788.7	519224.5	1.134
Ekiti	Rural	156148.0	143888.0	1.085
Enugu	Urban	194901.2	158688.6	1.228
Enugu	Rural	531911.2	449391.9	1.184
FCT[1]	Urban	198213.7	156628.1	1.266
FCT[1]	Rural	18689.1	15073.4	1.240
Gombe	Urban	132058.2	121877.1	1.084
Gombe	Rural	345265.4	325096.8	1.062
Imo	Urban	371848.8	327286.4	1.136
Imo	Rural	880127.7	796273.8	1.105
Jigawa	Urban	534233.5	486526.6	1.098
Jigawa	Rural	476047.3	431116.7	1.104
Kaduna	Urban	945498.5	847365.9	1.116
Kaduna	Rural	636545.6	601967.3	1.057
Kano	Urban	1267268.1	1056935.6	1.199
Kano	Rural	705160.0	596996.4	1.181
Katsina	Urban	369589.6	309068.0	1.196

Table 5: Weighted sums of design weight for all and completed households and nonresponse adjustment factors by state and urban/rural (continued)				
		Weighted sums of design weight		HH nonresponse adjustment factors
State	Urban/Rural	All households	Completed households	
Katsina	Rural	1763726.5	1596786.2	1.105
Kebbi	Urban	202622.3	172154.3	1.177
Kebbi	Rural	883777.7	782091.3	1.130
Kogi	Urban	615372.2	536047.6	1.148
Kogi	Rural	465580.0	413929.5	1.125
Kwara	Urban	560944.5	471410.1	1.190
Kwara	Rural	449981.5	366762.6	1.227
Lagos	Urban	1017011.6	712631.0	1.427
Lagos	Rural	67938.8	49953.3	1.360
Nasarawa	Urban	158105.5	137092.8	1.153
Nasarawa	Rural	316991.0	269543.8	1.176
Niger	Urban	288638.9	255889.9	1.128
Niger	Rural	1367087.4	1262700.8	1.083
Ogun	Urban	527451.5	388310.0	1.358
Ogun	Rural	284651.2	217390.3	1.309
Ondo	Urban	555860.7	475764.5	1.168
Ondo	Rural	719005.7	611277.6	1.176
Osun	Urban	1565024.3	1412907.6	1.108
Osun	Rural	244930.8	211432.0	1.158
Oyo	Urban	1564886.0	1425658.5	1.098
Oyo	Rural	608993.7	544230.8	1.119
Plateau	Urban	361313.6	326019.0	1.108
Plateau	Rural	760388.6	706603.1	1.076
Rivers	Urban	438129.4	348967.7	1.256
Rivers	Rural	775033.4	589111.1	1.316
Sokoto	Urban	284912.9	244541.4	1.165
Sokoto	Rural	603547.3	543885.7	1.110
Taraba	Urban	118244.0	111362.3	1.062
Taraba	Rural	581919.7	543307.6	1.071
Yobe	Urban	275444.3	239049.1	1.152
Yobe	Rural	801855.9	725068.1	1.106
Zamfara	Urban	366690.4	314444.5	1.166
Zamfara	Rural	557174.1	506722.9	1.100
[1] FCT: Federal Capital Territory				

The same process described above was used to calculate household weights for the pediatric and the Hepatitis subsamples, after adjusting the design weights using the subsample selection probabilities as indicated in Table 6. The nonresponse adjustment factors are presented in Table 7. The two subsample weights $hhwtp_{hij}$ and $hhwts_{hij}$ are not released in the final data files and are not to be used in data analysis. The two weights are calculated to facilitate the calculations of individual and blood weights for subsamples based on proper household survey weights, such as the adolescent interview weight *adowgt*, the blood draw weight *bdwgt* for children 0-14, and the Hepatitis weight *hepwgt*.

Table 6: Calculations of household weights for pediatric and hepatitis subsamples

Steps	Pediatric subsample	Hepatitis subsample
The subsample selection probabilities	$P_{3hi}^{-1} = g_{hi}^{p}/g_{hi}$ g_{hi}^{p} the number of households selected for the pediatric sample in cluster *i* in stratum *h*	$P_{4hi}^{-1} = g_{hi}^{s}/g_{hi}$ g_{hi}^{s} the number of households selected for the hepatitis sample in cluster *i* in stratum *h*
The nonresponse adjustment factor	$A_c^{HHp} = \sum_{j=1}^{n_c^p} P_{3hi}^{-1} D_{hij} \Big/ \sum_{j=1}^{n_c^p} R_{hij} P_{3hi}^{-1} D_{hij}$ n_c^p is the number of sampled pediatric households in adjustment cell *c*	$A_c^{HHs} = \sum_{j=1}^{n_c^s} P_{4hi}^{-1} D_{hij} \Big/ \sum_{j=1}^{n_c^s} R_{hij} P_{4hi}^{-1} D_{hij}$ n_c^s is the number of sampled Hepatitis households in adjustment cell *c*
The subsample household weight	$hhwtp_{hij} = A_c^{HHp} P_{3hi}^{-1} D_{hij}$	$hhwts_{hij} = A_c^{HHs} P_{4hi}^{-1} D_{hij}$

Table 7: Weighted sums of design weight for all and completed households and nonresponse adjustment factors by state and urban/rural

		Weighted sums of design weight				HH nonresponse adjustment factors	
		Pediatric subsample		Hepatitis subsample		Pediatric subsample	Hepatitis subsample
State	Urban/ Rural	All households	Completed households	All households	Completed households		
Abia	Urban	242212.9	209790.9	236009.2	204595.2	1.155	1.154
Abia	Rural	525325.5	481236.2	566565.0	543416.6	1.092	1.043
Adamawa	Urban	234700.9	206124.6	170179.8	153984.5	1.139	1.105
Adamawa	Rural	622040.7	565711.6	671624.3	590901.9	1.100	1.137
Akwa Ibom	Urban	179755.8	150145.6	145553.7	110514.9	1.197	1.317
Akwa Ibom	Rural	953671.0	856624.2	988987.7	895575.3	1.113	1.104
Anambra	Urban	893740.7	743451.1	922667.1	750274.1	1.202	1.230
Anambra	Rural	169439.2	144510.0	170495.6	129874.9	1.173	1.313
Bauchi	Urban	110936.0	88807.2	193075.9	158431.8	1.249	1.219
Bauchi	Rural	824181.3	767036.2	1108930.8	1029274.3	1.075	1.077
Bayelsa	Urban	129762.4	109904.7	127574.8	95817.2	1.181	1.331
Bayelsa	Rural	421272.0	366919.2	351825.7	303488.1	1.148	1.159

Table 7: Weighted sums of design weight for all and completed households and nonresponse adjustment factors by state and urban/rural (continued)							
		Weighted sums of design weight				HH nonresponse adjustment factors	
		Pediatric subsample		Hepatitis subsample		Pediatric subsample	Hepatitis subsample
State	Urban/ Rural	All households	Completed households	All households	Completed households		
Benue	Urban	181927.9	169734.0	138977.5	124189.1	1.072	1.119
Benue	Rural	1388014.2	1252751.7	1438982.8	1282661.9	1.108	1.122
Borno	Urban	641779.1	477420.9	650589.3	489214.2	1.344	1.330
Borno	Rural	273930.8	226573.2	278864.6	236197.5	1.209	1.181
Cross River	Urban	136535.3	109998.6	109262.9	87671.2	1.241	1.246
Cross River	Rural	662585.7	542843.4	431843.4	350300.4	1.221	1.233
Delta	Urban	452551.6	400174.8	492108.6	431441.9	1.131	1.141
Delta	Rural	733591.4	629227.0	754303.9	667683.4	1.166	1.130
Ebonyi	Urban	171641.4	161061.4	159639.1	146517.1	1.066	1.090
Ebonyi	Rural	755985.4	728368.6	847267.8	792712.3	1.038	1.069
Edo	Urban	536780.2	468335.8	446109.0	372212.0	1.146	1.199
Edo	Rural	322896.9	293414.3	404118.1	378280.3	1.100	1.068
Ekiti	Urban	595050.7	531440.0	625919.5	548947.7	1.120	1.140
Ekiti	Rural	141965.2	129922.0	154880.5	140873.5	1.093	1.099
Enugu	Urban	192340.9	153532.2	155532.5	116858.6	1.253	1.331
Enugu	Rural	524776.6	441806.3	414479.4	360148.5	1.188	1.151
FCT[1]	Urban	188554.1	144583.2	191371.2	151940.8	1.304	1.260
FCT[1]	Rural	16712.5	11079.5	19159.2	17417.2	1.508	1.100
Gombe	Urban	128066.9	115137.0	140584.3	138884.3	1.112	1.012
Gombe	Rural	330786.6	318162.5	368531.4	346769.9	1.040	1.063
Imo	Urban	386219.1	339340.3	391168.7	338061.1	1.138	1.157
Imo	Rural	866110.8	790618.7	890873.9	777012.3	1.095	1.147
Jigawa	Urban	521484.6	469031.8	534004.3	494766.2	1.112	1.079
Jigawa	Rural	465987.1	429831.8	478897.6	426316.0	1.084	1.123
Kaduna	Urban	948134.6	858462.0	876553.1	778582.5	1.104	1.126
Kaduna	Rural	604540.2	571237.6	660065.2	612030.9	1.058	1.078
Kano	Urban	1224728.8	1013186.8	3732562.4	3158625.4	1.209	1.182
Kano	Rural	703569.5	594815.5	2040182.1	1679028.7	1.183	1.215
Katsina	Urban	357269.4	283125.9	337735.3	291404.8	1.262	1.159
Katsina	Rural	1779004.3	1582123.7	1793989.8	1647092.0	1.124	1.089
Kebbi	Urban	215802.3	177230.5	212824.5	174180.1	1.218	1.222
Kebbi	Rural	900896.5	795493.2	866613.3	772881.1	1.133	1.121
Kogi	Urban	606161.8	536589.7	572297.3	523109.2	1.130	1.094

Table 7: Weighted sums of design weight for all and completed households and nonresponse adjustment factors by state and urban/rural (continued)							
		Weighted sums of design weight				HH nonresponse adjustment factors	
		Pediatric subsample		Hepatitis subsample		Pediatric subsample	Hepatitis subsample
State	Urban/ Rural	All households	Completed households	All households	Completed households		
Kogi	Rural	471407.6	413205.1	478654.7	430074.8	1.141	1.113
Kwara	Urban	565529.0	471211.3	573218.3	447049.4	1.200	1.282
Kwara	Rural	435644.4	341005.8	460395.9	370690.2	1.278	1.242
Lagos	Urban	1034392.9	728279.5	1243794.0	900547.2	1.420	1.381
Lagos	Rural	61832.5	44598.9	71545.7	50631.7	1.386	1.413
Nasarawa	Urban	152783.8	133841.3	98528.0	79755.5	1.142	1.235
Nasarawa	Rural	305575.0	253117.9	204819.5	174973.0	1.207	1.171
Niger	Urban	302121.2	270091.0	293015.7	258541.7	1.119	1.133
Niger	Rural	1369034.4	1250988.7	1279552.0	1163666.2	1.094	1.100
Ogun	Urban	498946.0	364526.0	576024.1	413499.1	1.369	1.393
Ogun	Rural	301191.1	221508.8	269207.3	205976.8	1.360	1.307
Ondo	Urban	582163.1	487150.5	558293.1	470316.9	1.195	1.187
Ondo	Rural	688152.2	581496.1	764864.3	626809.7	1.183	1.220
Osun	Urban	1582095.0	1411957.4	1640403.8	1425582.4	1.120	1.151
Osun	Rural	284451.8	243664.3	137562.2	131855.4	1.167	1.043
Oyo	Urban	1501199.3	1383867.0	1548574.5	1410865.9	1.085	1.098
Oyo	Rural	657057.8	574273.3	664546.3	601601.3	1.144	1.105
Plateau	Urban	340798.4	315830.8	390650.6	355793.7	1.079	1.098
Plateau	Rural	788311.0	734723.6	759318.7	716932.7	1.073	1.059
Rivers	Urban	429063.7	343176.8	470657.5	361254.8	1.250	1.303
Rivers	Rural	788816.8	607127.5	731362.7	540768.5	1.299	1.352
Sokoto	Urban	301751.2	249778.1	315949.1	267313.4	1.208	1.182
Sokoto	Rural	585729.7	521132.5	581190.1	520243.1	1.124	1.117
Taraba	Urban	108400.4	106984.5	122135.9	113069.4	1.013	1.080
Taraba	Rural	587743.2	552718.6	571935.1	531185.6	1.063	1.077
Yobe	Urban	267235.9	239122.3	327218.7	291790.7	1.118	1.121
Yobe	Rural	844082.7	755892.2	811757.1	748436.7	1.117	1.085
Zamfara	Urban	370006.1	305032.5	342655.2	304804.2	1.213	1.124
Zamfara	Rural	550029.8	503366.0	553844.0	499651.3	1.093	1.108
[1] FCT: Federal Capital Territory							

The household survey weights were post-stratified using the percentage distribution of the 2018 state-level population projections. The goal of this post-stratification was to determine the percentage distribution of households across states. The distribution of households at the state level would be the ideal to use for this adjustment, but these data were not available. The state-level population distribution was used instead, assuming a uniform household size across all states. The post-stratification factors are presented in Table 8.

Table 8: Projected and weighted households' distributions and post-stratification factors by states

	Households' distribution (%)				Post-stratification factors		
State	**Projected distribution**	**Weighted: all households (A)**	**Weighted: Pediatric households (B)**	**Weighted: Hepatitis households (C)**	**(A)**	**(B)**	**(C)**
Abia	1.99	1.9	1.92	1.82	1.047	1.036	1.093
Adamawa	1.94	2.16	2.14	1.91	0.896	0.905	1.014
Akwa Ibom	3.03	2.83	2.83	2.57	1.07	1.07	1.179
Anambra	3.19	2.64	2.66	2.48	1.21	1.201	1.288
Bauchi	3.04	2.4	2.34	2.95	1.268	1.3	1.031
Bayelsa	1.12	1.37	1.38	1.09	0.818	0.812	1.028
Benue	2.83	3.94	3.92	3.58	0.719	0.722	0.791
Borno	2.78	2.33	2.29	2.11	1.194	1.215	1.319
Cross River	2.14	1.98	2	1.23	1.081	1.07	1.74
Delta	3.19	3	2.96	2.83	1.064	1.078	1.128
Ebonyi	1.34	2.31	2.32	2.28	0.578	0.576	0.586
Edo	2.39	2.09	2.15	1.93	1.144	1.113	1.239
Ekiti	1.56	1.86	1.84	1.77	0.837	0.846	0.88
Enugu	2.49	1.81	1.79	1.29	1.376	1.392	1.931
FCT[1]	1.03	0.54	0.51	0.48	1.91	2.022	2.148
Gombe	1.27	1.19	1.15	1.15	1.066	1.103	1.103
Imo	2.96	3.12	3.13	2.91	0.947	0.944	1.015
Jigawa	2.87	2.52	2.47	2.3	1.138	1.161	1.247
Kaduna	4.29	3.94	3.88	3.48	1.089	1.106	1.233
Kano	6.24	4.92	4.82	13.09	1.268	1.295	0.477
Katsina	3.92	5.32	5.34	4.83	0.738	0.735	0.813
Kebbi	1.91	2.71	2.79	2.45	0.705	0.685	0.78
Kogi	2.14	2.69	2.69	2.38	0.796	0.796	0.9
Kwara	1.42	2.52	2.5	2.34	0.563	0.568	0.607
Lagos	9.15	2.7	2.74	2.98	3.39	3.34	3.071
Nasarawa	0.96	1.18	1.15	0.69	0.818	0.839	1.398
Niger	2.58	4.13	4.18	3.57	0.625	0.617	0.723
Ogun	3.08	2.02	2	1.92	1.522	1.538	1.602

Table 8: Projected and weighted households' distributions and post-stratification factors by states (continued)							
	Households' distribution (%)				Post-stratification factors		
State	Projected distribution	Weighted: all households (A)	Weighted: Pediatric households (B)	Weighted: Hepatitis households (C)	(A)	(B)	(C)
Ondo	2.62	3.18	3.17	3	0.823	0.825	0.872
Osun	2.52	4.51	4.66	4.03	0.558	0.54	0.625
Oyo	4.72	5.42	5.39	5.02	0.871	0.876	0.94
Plateau	2.04	2.8	2.82	2.61	0.73	0.725	0.783
Rivers	4.27	3.02	3.04	2.73	1.415	1.405	1.565
Sokoto	2.35	2.21	2.22	2.03	1.064	1.059	1.158
Taraba	1.31	1.75	1.74	1.57	0.747	0.752	0.833
Yobe	1.32	2.69	2.78	2.58	0.49	0.474	0.51
Zamfara	2.00	2.3	2.3	2.03	0.87	0.87	0.986
[1] FCT: Federal Capital Territory							

3.4.3 Adult Interview Weight (*adwgt*)

In completed households, all adults aged 15-64 years were eligible to complete the adult interview. Table 9 presents the number of eligible adults aged 15-64 years distributed over the response groups.

Table 9: Number of eligible adults aged 15-64 years by response groups		
Interview result	**Number of adults aged 15-64 years**	**Percentage (%)**
Completed	186,405	90.13
Not at home	9,369	4.53
Refused	8,515	4.12
Partly completed	1	0.00
Incapacitated	640	0.31
Others	1897	0.92
Total	206,827	100

The first step in the calculation of the adult interview weight was to adjust the household survey weight *hhwgt* for individual nonresponse, where adjustment factors inflated the *hhwgt* to account for the non-respondents. An adjustment factor A_k^I was calculated within each adjustment cell *c* as:

$$A_c^I = \sum_{k=1}^{m_c} hhwgt_{hijk} \Big/ \sum_{k=1}^{m_c} R_{hijk}^I hhwgt_{hijk}$$

where R_{hijk}^I =1 if the adult *k* in household *j* was respondent, and R_{hijk}^I =0 if s/he was a non-respondent, and m_c is the number of eligible adults in adjustment cell *c*. For the respondent adults, the nonresponse adjusted weight for adult *k* in household *j* in PSU *i* in stratum *h* was then computed as:

$adwgt_{hijk}^0 = A_c^I \; hhwgt_{hijk}$

The nonresponse adjustment cells *c* were determined through a two-stage process where 45 variables from the household questionnaire were used as covariates in a Least Absolute Shrinkage and Selection Operator (LASSO) model that was used to model the adult response to the individual questionnaire. The LASSO model's significant covariates (40 variables), along with gender, age, state and urban/rural, were inserted as inputs for a Chi-square Automatic Interaction Detector (CHAID) algorithm that identified the adjustment cells. A total of 105 adjustment cells were identified. For more details about the list of variables used, the LASSO/CHAID algorithms and the adjustment cells, the following materials are in theSampling and Weighting report_Attachments folder.

- VARs_for_ADRR.rtf: a list of variables used in the LASSO model.
- LASSO_CHAID_Adults.R: R code for LASSO model and CHAID algorithm.
- Adults tree.txt: R code for the CHAID algorithm and the tree.
- treeplot_adults.pdf: a plot for the CHAID tree.
- AD_NR_fact.csv: nonresponse adjustment factors by adjustment cells.

To reduce the variability of the weights which can lead to inflated sampling variances, the nonresponse adjusted weights $adwgt^{0}_{hijk}$ were trimmed where outliers (identified as greater than 3.5 times the median of the $adwgt^{0}_{hijk}$ within the corresponding sampling stratum) were capped at 3.5 times the median weight, yielding the trimmed nonresponse adjusted weight $adwgt^{0t}_{hijk}$. Finally, the trimmed weights were calibrated to the 2018 population projections of adults (aged 15-64 years) by gender and 10 age groups, yielding the final adult interview weight as follows:

$$adwgt_{hijk} = adwgt^{0t}_{hijk} \frac{M_c}{\sum_{k=1}^{m_c} R^{I}_{hijk} adwgt^{0t}_{hijk}}$$

where M_c is the 2018 projected population total in calibration cell *c*. Table 10 presents population projections, weighted totals, and calibration factors by age and gender.

Table 10: Projected and weighted totals of adults aged 15-64 years and calibration factors by gender and age

Age	Gender	2018 population projections	Weighted totals	Calibration factors
15 – 19 years	Male	10,530,755	8,360,266.04	1.26
15 – 19 years	Female	9,987,912	9,250,032.85	1.08
20 – 24 years	Male	8,719,592	6,074,567.78	1.44
20 – 24 years	Female	8,290,314	8,099,443.51	1.02
25 – 29 years	Male	7,365,365	6,062,248.65	1.21
25 – 29 years	Female	7,026,894	8,452,358.29	0.83
30 – 34 years	Male	6,401,791	5,263,979.90	1.22
30 – 34 years	Female	6,120,352	7,054,928.78	0.87
35 – 39 years	Male	5,535,047	5,070,010.01	1.09
35 – 39 years	Female	5,290,534	6,072,889.34	0.87
40 – 44 years	Male	4,528,562	4,087,418.46	1.11
40 – 44 years	Female	4,338,966	4,919,041.75	0.88
45 – 49 years	Male	3,481,557	3,351,177.89	1.04
45 – 49 years	Female	3,379,055	3,392,363.72	1.00

Table 10: Projected and weighted totals of adults aged 15-64 years and calibration factors by gender and age (continued)				
Age	**Gender**	**2018 population projections**	**Weighted totals**	**Calibration factors**
50 – 54 years	Male	2,781,726	2,889,092.40	0.96
50 – 54 years	Female	2,759,058	3,322,454.46	0.83
55 – 59 years	Male	2,147,661	2,141,102.48	1.00
55 – 59 years	Female	2,195,216	1,812,939.76	1.21
60 – 64 years	Male	1,617,987	2,401,347.05	0.67
60 – 64 years	Female	1,715,482	2,149,093.90	0.80

3.4.4 Adolescents Interview Weight (*adowgt*)

In the pediatric subsample, all adolescents (aged 10-14 years) were eligible to complete the adolescent interview. Table 11 presents the number of eligible adolescents aged 10-14 years distributed over the response groups.

Table 11: Number of eligible adolescents aged 10-14 years by response groups		
The individual interview result	**Number of adolescents aged 10-14 years**	**Percentage (%)**
Completed	10,665	85.83
Not at home	606	4.88
Refused	876	7.05
Incapacitated	30	0.24
Others	249	2.01
Total	12,426	100

The first step in the calculation of the adolescent interview weight was to adjust the pediatric household survey weight *hhwtp* for individual nonresponse, where adjustment factors inflated the *hhwtp* to account for the nonrespondents. Similar to the adjustment for nonresponse in the adult interview weight, the adjustment factor A_k^{Ia} was calculated for each adjustment cell *c* as:

$$A_c^{Ia} = \sum_{k=1}^{\acute{m}_c} hhwtp_{hijk} \Big/ \sum_{k=1}^{\acute{m}_c} R_{hijk}^{Ia} hhwtp_{hijk}$$

where R_{hijk}^{Ia} =1 if the adolescent *k* in household *j* was respondent, and R_{hijk}^{Ia} =0 if s/he was nonrespondent, and $\acute{m}_c$ is the number of eligible adolescents in adjustment cell *c*. For the respondent adolescents, the nonresponse adjusted weight for adolescent *k* in household *j* in PSU *i* in stratum *h* was then computed as:

$$adowgt_{hijk}^{0} = A_c^{Ia}\ hhwtp_{hijk}$$

Similar to the adult interview weight, the nonresponse adjustment cells *c* were determined through a two-stage process where 45 variables from the household questionnaire were used as covariates in a LASSO model that was used to model the adolescent's response to the individual questionnaire. The significant covariates (17 variables) along with gender, age, state and urban/rural were inserted as inputs for a CHAID algorithm that identified the adjustment cells. A total of 32 adjustment cells were identified.

For more details about the list of variables used, the LASSO/CHAID algorithms and the adjustment cells, the following materials are in the Sampling and Weighting report_Attachments folder:

- VARs_for_ADORR.rtf: a list of variables used in the LASSO model.
- LASSO_CHAID_ Adolescents.R: R code for LASSO model and CHAID algorithm.
- Adolescents tree.txt: R code for the CHAID algorithm and the tree.
- treeplot_adolescents.pdf: a plot for the CHAID tree.
- ADO_NR_fact.csv: nonresponse adjustment factors by adjustment cells.

Finally, after trimming the nonresponse adjusted weight $adowgt^{0}_{hijk}$ utilizing the same procedure for the adult interview weight, the trimmed weights $adowgt^{0t}_{hijk}$ were calibrated to the 2018 population projected totals of adolescents 10-14 years by gender, yielding the final adolescent's interview weight as follows:

$$adowgt_{hijk} = adowgt^{0t}_{hijk} \frac{M_c}{\sum_{k=1}^{m_c} R^{Ia}_{hijk} adowgt^{0t}_{hijk}}$$

where M_c is the 2018 projected population total in calibration cell *c*. Table 12 presents the population projections, weighted totals, and calibration factors by gender.

Table 12: Projected and weighted totals of adolescents aged 10-14 years and calibration factors by gender

Gender	2018 population projections	Weighted totals	Calibration factors
Male	12,416,800	10,673,045.97	1.16
Female	11,779,062	10,580,120.48	1.11

3.4.5 NSUM Weight (*nswgt*)

In all households, one adult (aged 18-54 years) was selected at random from each household to complete the NSUM module. The calculation of the NSUM weight is necessary to account for the within household selection probabilities. To calculate the NSUM weight, the adult interview weight *adwgt* for each adult (aged 18-54 years) *k* who completed the NSUM module was multiplied by the number of adults (18-54 years) within his/her household as follows:

$$nswgt^{0}_{hijk} = adwgt_{hijk} \times m^{(18-54)}_{hij}$$

where $m^{(18-54)}_{hij}$ is the number of adults (18-54 years) in household *j* in cluster *i* in stratum *h*. The $m^{(18-54)}_{hij}$ numbers were capped at 5 persons where $m^{(18-54)}_{hij}$ was coded as 5 in households with more than 5 adults aged 18-54 years. The NSUM weight $nswgt^{0}_{hijk}$ was then trimmed following the same trimming procedure used before and then the trimmed weight $nswgt^{0t}_{hijk}$ was calibrated to the 2018 population projections of adults (18-54 years) by gender and 8 age groups, yielding the final NSUM weight as follows:

$$nswgt_{hijk} = nswgt^{0t}_{hijk} \frac{M_c}{\sum_{k=1}^{m_c^{(18-54)}} nswgt^{0t}_{hijk}}$$

where M_c is the 2018 projected population total in calibration cell *c* and $m_c^{(18-54)}$ is the number of adults (18-54 years) who were selected for the NSUM module in calibration cell *c*. Table 13 presents the calibration factors by age and gender.

Table 13: Projected and weighted totals of adults aged 18-54 years and calibration factors by gender and age				
Age	**Gender**	**2018 population projections**	**Weighted totals**	**Calibration factors**
18 – 19 years	Male	4,212,302	3,348,619.92	1.26
18 – 19 years	Female	3,995,165	3,528,168.75	1.13
20 – 24 years	Male	8,719,592	7,032,884.00	1.24
20 – 24 years	Female	8,290,314	7,400,007.92	1.12
25 – 29 years	Male	7,365,365	6,468,964.35	1.14
25 – 29 years	Female	7,026,894	6,502,783.02	1.08
30 – 34 years	Male	6,401,791	5,786,469.50	1.11
30 – 34 years	Female	6,120,352	5,508,082.54	1.11
35 – 39 years	Male	5,535,047	5,116,769.78	1.08
35 – 39 years	Female	5,290,534	4,721,382.06	1.12
40 – 44 years	Male	4,528,562	4,288,208.38	1.06
40 – 44 years	Female	4,338,966	3,928,394.57	1.10
45 – 49 years	Male	3,481,557	3,434,196.09	1.01
45 – 49 years	Female	3,379,055	3,015,320.70	1.12
50 – 54 years	Male	2,781,726	2,621,557.89	1.06
50 – 54 years	Female	2,759,058	2,545,832.77	1.08

3.4.6 Blood Draw Weight (*bdwgt*)

All adults (aged 15-64 years) from all households and all children (aged 0-14 years) from the pediatric subsample households were eligible to have blood draws for an HIV test. Blood draw survey weights were calculated for adults (aged 15-64 years) and children (aged 0-14 years) separately and then concatenated in one variable *bdwgt*. Tables 14 and 15 presents the number of adults aged 15-64 years and children aged 0-14 years eligible for blood draws distributed over the response groups. Adults and children with blood draw that resulted in a valid HIV test (positive or negative HIV status) were coded as completed. If eligible for the blood draw, adults and children who did not consent for the blood draw or cases with invalid HIV tests were coded as not completed.

Table 14: Number of eligible adults aged 15-64 years by response groups		
Blood draw result	**Number of adults aged 15-64 years**	**Percentage (%)**
Completed	173,716	93.19
Not completed	12,689	6.81
Total	186,405	100

Table 15: Number of eligible children aged 0-14 years by response groups		
Blood draw result	**Number of children aged 0-14 years**	**Percentage (%)**
Completed	32,494	71.14
Not completed	13,179	28.86
Total	45,673	100

The first step in the process of calculating the blood draw weights was adjusting the relevant survey weight for blood draw nonresponse; i.e., the adults survey weight *adwgt* for adults aged 15-64 years, and the pediatric household survey weight *hhwtp* for children aged 0-14 years. For adults aged 15-64 years, determining the adjustment cells through the LASSO/CHAID approach was done separately for males and females, with variables from the individual questionnaire added to variables from the household questionnaire and utilized in the LASSO model. Table 16 presents the number of variables used in the LASSO model and CHAID algorithm and the number of nonresponse adjustment cells for adults and children.

Table 16: Number of variables used in LASSO/CHAID and number of nonresponse adjustment cells

Age group	Gender	LASSO	CHAID	Adjustment cells
Adults 15-64 years	Male	60	47	44
Adults 15-64 years	Female	63	60	54
Children 0-14 years	All	44	40	51

For more details, Table 17 presents the relevant materials that are in the Sampling and Weighting report_Attachments folder.

Table 17: Number of eligible children 0-14 by response groups

Adults 15-64 years	Children 0-14 years	Description
VARs_for_BADRR.rtf	VARs_for_BCHRR.rtf	A list of variables used in the LASSO model.
LASSO_CHAID_AD Blood.R	LASSO_CHAID_CH Blood.R	R codes for LASSO model and CHAID algorithm.
Adblood_males_tree.txt Adblood_females_tree.txt	Chblood_tree.txt	R codes for the CHAID algorithm and the tree.
treeplot_adblood_Males.pdf treeplot_adblood_Females.pdf	treeplot_chblood.pdf	A plot for the CHAID tree.
ADB_NR_fact.csv	CHB_NR_fact.csv	Nonresponse adjustment factors by adjustment cells.

After adjusting for nonresponse, the concatenated weights were trimmed and calibrated to the relevant projected population totals. Table 18 presents the calibration factors by gender and age.

Table 18: Projected and weighted totals of persons 0-64 and calibration factors by gender and age

Age	Gender	2018 population projections	Weighted totals	Calibration factors
0–4 years	Male	16,350,611	17,538,222.75	0.932
0–4 years	Female	15,561,184	16,263,205.22	0.957
5–9 years	Male	14,370,262	14,454,085.54	0.994
5–9 years	Female	13,652,524	14,167,804.94	0.964
10–14 years	Male	12,416,800	7,991,494.13	1.554
10–14 years	Female	11,779,062	7,760,614.18	1.518
15–19 years	Male	10,530,755	10,530,408.42	1.000

Table 18: Projected and weighted totals of persons 0-64 and calibration factors by gender and age (continued)				
Age	**Gender**	**2018 population projections**	**Weighted totals**	**Calibration factors**
15–19 years	Female	9,987,912	9,989,666.19	1.000
20–24 years	Male	8,719,592	8,658,834.44	1.007
20–24 years	Female	8,290,314	8,306,920.62	0.998
25–29 years	Male	7,365,365	7,352,025.21	1.002
25–29 years	Female	7,026,894	7,055,001.57	0.996
30–34 years	Male	6,401,791	6,396,746.86	1.001
30–34 years	Female	6,120,352	6,126,675.34	0.999
35–39 years	Male	5,535,047	5,552,979.40	0.997
35–39 years	Female	5,290,534	5,313,825.32	0.996
40–44 years	Male	4,528,562	4,547,308.08	0.996
40–44 years	Female	4,338,966	4,344,224.09	0.999
45–49 years	Male	3,481,557	3,484,928.56	0.999
45–49 years	Female	3,379,055	3,365,475.05	1.004
50–54 years	Male	2,781,726	2,778,818.90	1.001
50–54 years	Female	2,759,058	2,738,416.08	1.008
55–59 years	Male	2,147,661	2,139,962.66	1.004
55–59 years	Female	2,195,216	2,148,477.96	1.022
60–64 years	Male	1,617,987	1,624,747.13	0.996
60–64 years	Female	1,715,482	1,713,325.08	1.001

3.4.7 Hepatitis Weight (*hepwgt*)

In the Hepatitis subsample households, one adult (aged 15-64 years) was selected at random from each household to be tested for Hepatitis B and C. All adults (aged 15-64 years) who tested positive for HIV were tested for Hepatitis. For the calculation of the Hepatitis weight for persons in the randomly selected subsample of households, it is necessary to account for the selection probabilities of selecting the subsample of households and for the within household selection probabilities. The weighting process for the subsample starts with the blood weight *bdwgt*, which was adjusted to account for the subsampling selection probabilities and for within household selection as follows:

$$hepwgt^{0}_{hijk} = bdwgt_{hijk} \times \frac{hhwts_{hijk}}{hhwgt_{hijk}} \times m^{(15-64)}_{hij}$$

where $m^{(15-64)}_{hij}$ is the number of adults (15-64 years) in household *j* in cluster *i* in stratum *h*. The $m^{(15-64)}_{hij}$ numbers were capped at 5 persons where $m^{(15-64)}_{hij}$ was coded as 5 in households with more than 5 adults age 15-64 years. The observations with the adjusted weights as described above were then concatenated with observations for the adults (15-64 years) who tested positive for HIV. For those persons who tested positive for HIV the *hepwgt* was set equal to the blood weight *bdwgt*. The *hepwgt's* on this concatenated file were then trimmed and calibrated to the relevant projected population totals. Table 19 presents the calibration factors by the calibration variables.

Table 19: Projected and weighted totals of persons 15-64 and calibration factors by gender and age				
Age	**Gender**	**2018 population projections**	**Weighted totals**	**Calibration factors**
15–19 years	Male	10,530,755	9,143,554.24	1.152
15–19 years	Female	9,987,912	9,760,673.21	1.023
20–24 years	Male	8,719,592	8,528,485.33	1.022
20–24 years	Female	8,290,314	8,475,555.31	0.978
25–29 years	Male	7,365,365	7,149,333.26	1.030
25–29 years	Female	7,026,894	7,008,445.37	1.003
30–34 years	Male	6,401,791	6,985,700.22	0.916
30–34 years	Female	6,120,352	6,016,154.74	1.017
35–39 years	Male	5,535,047	5,395,078.05	1.026
35–39 years	Female	5,290,534	5,139,616.48	1.029
40–44 years	Male	4,528,562	4,695,606.05	0.964
40–44 years	Female	4,338,966	4,448,794.98	0.975
45–49 years	Male	3,481,557	3,693,641.92	0.943
45–49 years	Female	3,379,055	3,315,919.25	1.019
50–54 years	Male	2,781,726	3,221,895.41	0.863
50–54 years	Female	2,759,058	2,713,807.02	1.017
55–59 years	Male	2,147,661	2,226,804.59	0.964
55–59 years	Female	2,195,216	2,373,530.29	0.925
60–64 years	Male	1,617,987	1,912,065.28	0.846
60–64 years	Female	1,715,482	1,980,787.32	0.866

3.5 Analysis and Variance Estimation

For data analysis, the appropriate weights for the specific analysis of interest should be utilized, which is generally determined by the target population of inference. Below are some guidelines regarding the different weights:

- Household weight *hhwgt* can be used for analyses conducted at the household level, for example, the distribution of households by urban/rural residence. The household weight can be interpreted as the number of households that the participating household represents in the population, accounting for sampling selection and non-response at the EA and household levels.
- Interview individual weights *adwgt* and *adowgt* can be used for analyses conducted at the individual level for data collected for all potentially eligible interview participants. For example, self-reported HIV testing (i.e., ever received an HIV test prior to the survey) should be estimated using interview weights since all interview respondents received HIV testing questions. In this scenario, interview weights can be interpreted as the number of individuals that the respondent represents in the population who could have participated in the interview, accounting for sampling and non-response at the EA, household and individual levels.
- Blood weight *bdwgt* can be used for analyses conducted only among blood test participants. For example, HIV prevalence should be estimated using blood test weights even if the analysis includes predictors at the household or individual level since not all interview

respondents participated in blood tests. In this scenario, each participant's blood weight can be interpreted as the number of individuals that the participant represents in the population who could have participated in blood testing, accounting for selection and non-response of EA, household, individual and blood testing. In addition, if the outcome of interest comes from the interview (e.g., HIV testing history), but the analysis is restricted to those who have blood test results, blood test weights should be used.

- Hepatitis weight *hepwgt* can be used for analysis conducted only among Hepatitis B and C test participants.

Multiple existing variance estimation methods can appropriately be used to estimate design-based standard errors for this complex survey. These are Taylor series linearization and replication approach such as Jackknife repeated replication method. These methods require specifying appropriate survey weights, strata (state) and primary sampling units (cluster). The NAIIS dataset includes identifier variables for sampling design strata, primary sampling unit or cluster, and survey weights. Users will need to specify these three variables for the analysis of interest at the national level.

Unlike the other PHIA surveys, Jackknife replicates are not being released with the final datasets due to a large number of the survey clusters. For users who are interested in estimating the variance using the Jackknife repeated replication method, SAS can be used for variance estimation with replicates created based on the survey clusters and strata.